

## HT1220 Report

---

### RECORD INFORMATION

Gene ID:	1220
Sequence ID:	1220
Protein ID:	1220
Sequence name:	thrombospondin 1, alt. transcript 1
Genome:	nucleus
Taxon:	Homo sapiens
Locus:	1220
Common Name:	thrombospondin 1
Role ID:	40

Coding sequence length:	3513 nt
Transcript sequence length:	5722 nt
Expression data:	<u>THC201673</u>

---

### ACCESSION DATA

HT1220 is derived from accessions(s):

SP:P07996 (THROMBOSPONDIN 1 PRECURSOR.)  
GB:X04665 (Human mRNA for thrombospondin)  
GB:X14787 (Human mRNA for thrombospondin)  
GB:U12471 (thrombospondin-p50 {Homo sapiens})  
GB:M99425 (Human thrombospondin mRNA, 3' end.)  
PIR:G01478 (thrombospondin-p50 - human (fragment))  
GB:U12471 (Human thrombospondin-1 gene, partial cds.)  
GB:J04835 (Human thrombospondin gene, exons 1, 2 and 3.)  
GB:M25631 (Homo sapiens (clone lambda-TS-33) thrombospondin (THBS) mRNA, 5' end.)

---

### ALTERNATIVE SPLICE INFORMATION

Alternative splice forms for this gene:

HT3987 thrombospondin 1, alt. transcript 2

---

### MAPPING DATA

GDB accession(s) for this gene:

GDB ID:	Symbol
-----	

Figure 1A

gdb:120438 THBS1

## cDNA FEATURES

Feature	End 5	End 3
-----	-----	-----
coding_seq	112	3624
3'UT	3625	5722
spjunc_h	1235	1236

## SEQUENCE

### nucleotide:

```

ggacgcacaggcattccccgcgcccctccagccctcgccgccctcgccaccgctccccggc
cgccgcgctccgggtacacacaggatccctgctgggcaccaacagctccaccatggggctg
gcctggggactaggcgtcctgttccctgatgcattgtgtgtggcaccacacgcattccagag
tctggcggagagacaacagcgtgtttgacatccttgaactcaccggggcgcccgcaagggg
tctggggcgccgactggtgaaggggccccgaccctccagcccagctttccgcacagaggat
gccaacctgatccccctgtgcctgatgacaagtccaagacctggtggatgctgtgccc
gcagaaaagggtttcctccttctggcatccctgaggcagatgaagaagaccggggcagc
ctgctggccctggagcggaaagaccactctggccaggtcttcagcgtggtgtccaatggc
aaggcggggcaccctggacctcagcctgaccgtccaaggaaagcagcacgtggtgtctgtg
gaagaagctctcctggcaaccggccagtggaagagcatcaccctgtttgtgcaggaagac
agggcccagctgtacatcgactgtgaaaagatggagaatgctgagttggacgtccccatc
caaagcgtcttcaccagagacctggccagcatcgccagactccgcacgcaaaaggggggc
gtcaatgacaatttccagggggtgctgcagaatgtgaggtttgtctttggaaccacacca
gaagacatcctcaggaacaaaaggctgctccagctctaccagtgctcctcacccttgac
aacaacgtggtgaaatggttccagccctgccatccgcactaactacattggccacaagaca
aaggacttgcaagccatctgcggcatctcctgtgatgagctgtccagcatggtcctggaa
ctcaggggctgcgcaccattgtgaccacgctgcaggacagcatccgcaaagtgactgaa
gagaacaaagagttggccaatgagctgaggcggcctccccctatgctatcacaacggagtt
cagtacagaaaataacgaggaatggactgttgatagctgcactgagtgctactgtcagaac
tcagttaccatctgcaaaaagggtgtcctgcccccatcatgcccgtgctccaatgccacagtt
cctgatggagaatgctgtcctcgtctgttggcccagcgactctgcggacgatggctgggtct
ccatggtccgagtggaacctcctgttctacgagctgtggcaatggaattcagcagcgccggc
cgctcctgcgatagcctcaacaaccgatgtgagggctcctcggtccagacacggacctgc
cacattcaggagtggtgacaaaagatttaaacaggatggtggctggagccactggtccccg
tggtcatcttgttctgtgacatgtggtgatggtgtgatcacaaggatccggctctgcaac
tctcccagccccagatgaatgggaaaacctgtgaaggcgaagcgcgggagaccaaaagcc
tgcaagaaagacgcctgcccccatcaatggaggctggggctccttggtcaccatgggacatc
tggtctgtcacctgtggaggaggggtacagaaacgtagtctctctgcaacaacccccgca
ccccagtttggaggcaaggactgcgttggtgatgtaacagaaaaccagatctgcaacaag
caggactgtccaattgatggatgcctgtccaatccctgctttgcggcgctgaagtgtact
agctaccctgatggcagctggaaaatgtggtgcttgtccccctgggttacagtggaaatggc
atccagtgacagatgttgatgagtgcaagaagtgcctgatgcctgcttcaaccacaat
ggagagcacccggtgtgagaacacggacccccggctacaactgcctgccttccccccacgc
ttcaccggctcacagcccttcggccagggtgtcgaaatgccaacggccaacaaacaggtg
tgcaagccccgtaacccctgcacggatgggacccacgactgcaacaagaacgcgaagtgc
aactacctggggccactatagcgaccccatgtaccgctgcgagtgcaagcctggctacgct
ggcaatggcatcatctgcggggaggacacagacctggatgggtggcccaatgagaacctg
gtgtgctggcccaatgcgacttaccactgcaaaaaggataattgccccaaccttcccaac
tcagggcaggaagactatgacaaggatggaattgggtgatgcctgtgatgatgacgatgac
aatgatataaattccagatgacagggacaactgtccattccattacaaccagctcagtat
gactatgacagagatgatgtgggagaccgctgtgacaactgtccctacaaccacaaccca

```

Figure 1B

gatcaggcagacacagacaacaatggggaaggagacgcctgtgctgcagacattgatgga  
gacgggtatcctcaatgaacgggacaactgccagtacgtctacaatgtggaccagagagac  
actgatatggatgggggttgagatcagtggtgacaattgccccttggaacacaatccggat  
cagctggactctgactcagaccgcattggagatacctgtgacaacaatcaggatattgat  
gaagatggccaccagaacaatctggacaactgtccctatgtgccaatgccaaccaggct  
gaccatgacaaaagatggcaaggagatgcctgtgaccacgatgatgacaacgatggcatt  
cctgatgacaaggacaactgcagactcgtgccaatcccagaccagaaggactctgacggc  
gatggtcgaggtgatgcctgcaaagatgattttgaccatgacagtggtgccagacatcgat  
gacatctgtcctgagaatgttgacatcagtgagaccgatttccgccgattccagatgatt  
cctctggaccccaaaggacatcccaaatgaccctaactgggttgtagcccatcagggt  
aaagaactcgtccagactgtcaactgtgatcctggactcgtgtaggttatgatgagttt  
aatgctgtggacttcagtggcaccttcttcatcaacaccgaaaaggacgatgactatgct  
ggattttgtctttggctaccagtcacagcagccgcttttatgttgtgatgtggaagcaagtc  
accagtcctactgggacaccaacccacaggggctcagggatactcgggcctttctgtg  
aaagttgtataaactccaccacagggcctggcgagcacctgcggaacgcctgtgggcacaca  
ggaaaacacccctggccaggtgcgacccctgtggcatgaccctcgtcacataggctggaaa  
gatttccaccgcctacagatggcgtctcagccacaggccaaagacgggtttcattagagt  
gtgatgtatgaagggaagaaaatcatggctgactcaggacccatctatgataaaacctat  
gctgggtggtagactagggttgtttgtcttctctcaagaaatgggtgttcttctctgacctg  
aaatacgaatgtagagatccctaatacatcaaatgttgattgaaagactgatcataaacc  
aatgctggtattgcaccttctggaactatgggcttgagaaaaacccaggatcacttctc  
cttggcttcttcttcttctgtgcttgcatcagtggtgactcctagaacgtgcgacctgcc  
tcaagaaaatgcagttttcaaaaacagactcatcagcattcagcctccaatgaataagac  
atcttccaagcatataaaacaattgctttgggttcttctttgaaaaagcatctacttgcttc  
agttgggaagggtgccattccactctgcctttgtcacagagcaggggtgctatttgtgaggc  
catctctgagcagtggtgactcaaaagcattttcaggcatgtcagagaaggaggactcact  
agaattagcaaaacaaaaccacctgacatcctccttcaggaacacggggagcagaggcca  
aagcactaagggggagggcgcataccggagacgattgtatgaagaaaatatggaggaactg  
ttacatgttccgtactaagtcattttcaggggattgaaagactattgctggatttcattga  
tgctgactggcggttagctgattaacccatgtaaataggcacttaaatagaagcaggaag  
ggagacaaaagactggcttctggacttccctcctgatccccacccttactcatcacttgc  
agtggccagaattaggggaatcagaatcaaaaccagtgtaaggcagtgctggctgccattgc  
ctgggtcacattgaaattgggtggcttcatttctagatgtagcttgtgacagatgtagcaggaa  
aataggaaaaacctaccatctcagtgagcaccagctgcctcccaaaggaggggcagccgtg  
cttatattttttatgggttacaatggcacaaaattattatcaacctaaactaaaacattcctt  
ttctcttttttccgtaattactaggtagttttctaatctctcttttggaaagtatgattt  
ttttaaagtctttacgatgtaaaatatttattttttacttattctggaagatctggctga  
aggattattcatggaacaggaagaagcgtaaagactatccatgtcatctttgttgagagt  
cttcgtgactgtaagattgtaaatacagattatttatttaactctgttctgctggaaatt  
taggcttcatacggaaaagtgtttgagagcaagtagttgacatttatcagcaaatctcttg  
caagaacagcacaaaggaaaatcagtcataaagctgctctgccccttgctcagagtgg  
atggttatgggattccttttttctgttttatcttttcaagtggaaattagttgggttatcc  
atttgcaaatgttttaaatgcaaaagccatgaggtcttcaatactgttttacccca  
tcccttgatgcatatttccaggggagaaggaaagcatatacacttttttcttcattttcc  
aaaagagaaaaaaatgacaaaagggtgaaacttacatacaaatattacctcatttgtgtg  
tgactgagtaaaagaatttttggatcaagcggaaaagagtttaagtgtctaaacaaactaaa  
gctactgtagtacctaaaaagtcagtggtgtacatagcataaaaaactctgcagagaagta  
ttcccaataaggaaatagcattgaaatgttaaatacaatttctgaaagttatgtttttt  
tctatcatctgggtataccattgctttatttttataaattattttctcattgccattggaa  
tagaatattcagatttgtgtagatatgctattttaaataatttatcaggaataactgcctgt  
agagttagttttctattttttatataatgtttgcacactgaattgaagaattgttgggtt  
tttctttttttgtttttttttttttttttttttttttttttttgttttgacctccattttta  
ctatttgccaatacctttttctaggaatgtgctttttttgtacacatttttatccattt  
tacattctaaagcagtgtaagttgtatattactgtttctttatgtacaaggaacaacaata  
aatcatatggaaaatttatattt

protein:

MGLAWGLGVFLMHVCGTNRIPESGGDNSVFDIFELTGAARKSGRRLVKGPDPSSPAFR

Figure 1C

IEDANLIPPVPDDKFQDLVDAVRAEKGFLLLASLRQMCKTRGTLALERKDHSGQVFSVV  
SNGKAGTLDLSLTVQKQHVVSVEEALLATGQWKSITLFFVQEDRAQLYIDCEKMENAEILD  
VPIQSVFTRDLASLARLRIAKGGVNDNFQGV LQNVRFVFGTTPEDILRNKGCSSSTSLL  
TLDNNVVGSSPAIRNTNYIGHKTKDLQAIICGISDELSSMVLELRGLRTIVTTLQDSIRK  
VTEENKELANELRRPPLCYHNGVQYRNNEEWTVDSCTECHCQNSVTICKKVSCPIMPCCSN  
ATVPDGECCPRCWPSPDSADDGWS PWSEWTSCSTSCGNGIQQRGRSCDSLNNRCEGSSVQT  
RTCHIQECDKRFKQDGGWSHWSPWSSCSVTCDGVITRIRLCNSPSPQMNGKPCGEARE  
TKACKKDACPINGGWGPWSPWDICSVTCGGGVQKRSRLCNPAPQFGGKDCVGDVTENQI  
CNKQDCPIDGCLSNPCFAGVKCTSYPDGSWKCGACPPGYSGNGIQCTDVDECKEVPDACF  
NHNGEHRCENTDPGYNCLPCPPRFTGSQPFQGVGHATANKQVCKPRNPCTDGTDCNKN  
AKCNYLGHYS DPMYRCECKPGYAGNGIICGEDTDLGWPNEENLVCVANATYHCKKONCPN  
LPNSGQEDYDKDIGDADDDDDNDKI PDDRDNC PFHYNPAQYDYDRDDVGDRCDCNCPYN  
HNPDQADTDNNGEGDACAADIDGDGILNERDNCQYVYNVDQRDTMDGVDGQCDNCPLEH  
NPDQLDSDSDRIGDTCNNQDIDEDGHQNNLDNCPYVNPANQADHDKDGKGDACDHDDDN  
DGIPDDKONCRLVPNPDQKSDGDGRGDACKDDFDHDSVPDIDDICPENVDISETDFRRF  
QMIPLDPKGTSQNDPNWVVRHQKELVQTVNCDPGLAVGYDEFNAVD FSGTFFINTERDD  
DYAGFVFGYQSSSRFYVVMWKQVTQSYWDTNPTRAQGYSGLSVKVNSTTGPGEHLRNAL  
WHTGNTPGQVRTLWHDPRHIGWKDFTAYRWRLSHRPKTGFIRVVMYEGKKIMADSGPIYD  
KTYAGGRLGLFVFSQEMVFFSDLKYECRDP

Figure 1D

## HT2143 Report

---

### RECORD INFORMATION

Gene ID: 2081  
Sequence ID: 2143  
Protein ID: 2125  
Sequence name: thrombospondin 4  
Genome: nucleus  
Taxon: Homo sapiens  
Locus: 2081  
Common Name: thrombospondin 4  
Role ID: 40

Coding sequence length: 2886 nt  
Transcript sequence length: 3074 nt  
Expression data: THC168897

---

### ACCESSION DATA

HT2143 is derived from accessions(s):

SP:P35443 (THROMBOSPONDIN 4 PRECURSOR.)  
GB:Z19585 (thrombospondin-4 {Homo sapiens})  
GB:Z19585 (H.sapiens mRNA for thrombospondin-4)  
PIR:A55710 (thrombospondin 4 precursor - human)

---

### cDNA FEATURES

Feature	End 5	End 3
coding_seq	28	2913
3'UT	2914	3074

---

### SEQUENCE

nucleotide:

gaattccggggagcaggaagagccaacatgctggccccgcgaggagccgcccgtcctcctg  
ctgcacctgggtcctgcagcgggtgctagcggcaggcgcccaggccacccccagggtctt  
gaccttctcccatcttccagtcagaggctaaacccaggcgctctgctgccagtcttgaca  
gaccccgccctgaatgatctctatgtgatttccaccttcaagctgcagactaaaagttca  
gccaccatcttcgggtctttactcttcaactgacaacagtaaatattttgaatttactgtg  
atgggacgcttaagcaaagccatcctccgttacctgaagaacgatgggaaggtgcatttg

Figure 2A

gtgggttttcaacaacctgcagctggcagacggaaggcgccacaggatcctcctgaggctg  
agcaatttgcagcgaggggcccggctccctagagctctacctggactgcatccagggtggat  
tccgttcacaatctccccagggcctttgctggccctcccagaaacctgagaccattgaa  
ttgaggactttccagaggaagccacaggacttcttggaagagctgaagctgggtgga  
ggctcactgttccagggtggccagcctgcaagactgcttctgcagcagagtgcaccactg  
gctgccacaggcacaggggactttaaccggcagttcttgggtcaaataacacaaat  
caactcctgggagaggtgaaggaccttctgagacagcagggttaaggaaacatcatttttg  
cgaaacaccatagctgaatgccaggcttgccgtcctctcaagtctcagctccgacccca  
agcacgggtgggtcgccccggctccccctgcaccgccaacacgcccacctcgctcggtgtgac  
tccaacccatgtttccgaggtgtccaatgtaccgacagtagagatggcttccagtgtggg  
ccctgccccgagggctacacaggaaacgggatcacctgtattgatgtgatgagtgcaaa  
taccatccctgctaccgggctgcaactgcataaatttgcctcctggcttcagatgtgac  
gcttgcccagtggtgttccagggcccatgggtgcagggtgttgggatcagtttggccaa  
tcaaacaagcagggtctgcactgacattgatgagtgctgaaatggagcgtgcgttcccaac  
tcgatctgcgttaatactttgggatcttaccgctgtgggccttgaagccgggggtatact  
gggtgatcagataaggggatgcaaagtggaaagaaactgcagaaacccagagctgaacct  
tgcagtgatgaatgccagtgcatgaagagaggcagggggatgtgacatgtgtgtgtgga  
gtcgggtgggctggagatggctatatctgtggaaaggatgtggacatcgacagttacccc  
gacgaagaactgccatgctctgccaggaactgtaaaaaggacaactgcaaatatgtgcca  
aatcttgcccaagaagatgcagacagagatggcatggcgacgcttgtgacgaggtgct  
gacggagatgggatcctgaatgagcaggataaactgtgtcctgattcataatgtggacca  
aggaacagcgataaagatatctttggggatgcctgtgataactgcctgagtgctttaa  
aacgaccagaaagacaccgatggggatggaagaggagatgcctgtgatgatgacatggat  
ggagatggaataaaaaaacattctggacaactgccccaaatctcccaatcgtgaccaacgg  
gacaaggatgggtgatgggtgtgggggatgcctgtgacagttgtcctgatgtcagcaacct  
aaccagtgctgatgtggataatgatctggttggggactcctgtgacaccaatcaggacag  
gatggagatgggacaccagagcagcagacaactgccccaccgtcatgaacagtgccacg  
ctggacaccgataaaggatggaattgggtgacgagtgatgatgatgatgacaatgatggt  
atcccagacctggtgccccctggaccagacaactgcccggctggtccccaaacccagcccag  
gaggatagcaacagcgacggagtgaggagacatctgtgagtgctgactttgaccaggaccag  
gtcatcgatcggtcgacgtctgcccagagaacgcagaggtcaccttgaccgacttcagg  
gcttaccagaccgtgggcctggatcctgaaggggatgcccagatcgatcccaactgggtg  
gtcctgaaccagggcatggagattgtacagaccatgaacagtgatcctggcctggcagtg  
gggtacacagcttttaattggagttgacttcgaaggaccttccatgtgaataccagaca  
gatgatgactatgcaggctttatctttgggtaccaagatagctccagcttctacgtggct  
atgtggaagcagacggagcagacatatgggaagccacccattccgagcagttgcagaa  
cctggcattcagctcaaggctgtgaagtctaagacaggtccaggggagcatctccggaac  
tccctgtggcacacgggggacaccagtgcaccaggtcaggctgctgtggaaggactccagg  
aatgtgggctggaaggacaagggtgtcctaccgctggttctacagcacaggccccagggtg  
ggctacatcagggtacgattttatgaaggctctgagttgggtgctgactctggcgtcacc  
atagacaccacaatgcgtggaggccgacttggcgttttctgcttctcaagaaaacatc  
atctgggtccaaacctcaagtatcgctgcaatgacaccatccctgaggacttccaagagttt  
caaaccacagaatttcgaccgcttcgataattaaaccaaggaagcaatctgtaactgcttt  
tcggaacactaaaaccatatatattttaacttcaattttcttttagcttttaccacccaa  
atatatcaaaacgttttatgtgaatgtggcaataaaggagaagagatcatttttaaaaaa  
aaaaaaaaaaaaaa

**protein:**

MLAPRGAAVLLHLVLQRWLAAGAQAATPQVFDLLPSSSQRLNPGALLPVLTDPALNDLYV  
ISTFKLQTKSSATIFGLYSTDNSKYFEFTVMGRLSKAILRYLKNDBGKVLVVFNNLQLA  
DGRHRILLRLSNLQRGAGSLELYLDCTIQVDSVHNLPRAFAGPSQKPEIETLRTFQRKPQ  
DFLEELKLVVRGSLFQVASLQDCFLLQSEPLAATGTGDFNRQFLGQMTQLNQLLGEVKDL  
LRQQVKETSFLRNTIAECQACGPKLFQSPTSTVVAPAPPAPTRPPRRCDNSNPCFRGVQ  
CTDSRDGFQCGPCPEGYTGNITCIDVDECKYHPCYPGVHCINLSPGFRCDACPVGFTGP  
MVQGVGISFAKSNKQVCTDIDECRNGACVPNSICVNTLGSYRCGPCKPGYTGDQIRGCKV  
ERNCRNPENLPCSVNAQCIEERQGDVTCVCGVGWAGDGYICGKDVIDSYDPDEELPC  
NCKKDNCYVPNSGQEDADRDGIGDACDEADGDGILNEQDNCVLIHNVQDQNSDKDIFG  
DACDNCLSVLNNQKDTGDGRGDACDDMDGDGINKILDNCPKFPNRDQRDKDGDGVGD

**Figure 2B**

ACDSCP DVSNPNQSDVDNDLVGDS CDTNQDS DGDGHQDSTDNCPTVINS AQLD TDKDGIG  
DECDDDDNDGIPDLVPPGPDNCR LVPNPAQEDSNSDGVGDICESDFDQDQVIDRIDVCP  
ENAEVTLTDFRAYQTVGLDPEGDAQIDPNWVVLNQGMEIVQTMNSDPGLAVGYTAFNGVD  
FEGTFHVNTQTDDDYAGFI FGYQDSSSFYVVMWKQTEQTYWQATPFRAVAEPGIQLKAVK  
SKTGPGEHLRNSLWHTGDTSDQVRL LWKDSRNVGWKDKVSYRWFLQHRPQVGYIRVRFYE  
GSELVADSGVTIDTTMRGGRLGVFCFSQENI IWSNLKYRCNDTIPEDFQEFQTQNFDRFD  
N

Figure 2C

Poly ID	Sequence ID	Pos Mon	Gene Description	Flanking Seq	Mutation T/Ref/NT	Ref AA	Alt AA
G334i4	11T111220_mRNA	2210	THBS1, thrombospondin 1	TCGATGGCTGGCCCAJA/GITGAGAACCTGGTGTG	Missense A	N	S
G395i2	11T112143_mRNA	1188	THBS4, thrombospondin 4	GAGTGTGGAATBGA/GC/CGTGC GTTCCCAACT	Missense G	A	P

Figure 3